GAU-2838
#18/Supp. Decl.
Shaw
2/3/00

OIP JC76
JAN 18 2000
PATENT & TRADEMARK OFFICE

Docket No. 18036-20
PATENT

I hereby certify that this correspondence is being deposited on **January 13, 2000**, with the United States Postal Service, first class mail, in an envelope addressed to: Box CPA, Assistant Commissioner for Patents, Washington, D.C. 20231

Jessica S. Brown
Jessica S. Brown

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of
Lars G. SVENSSON et al.

Title: SYSTEM AND METHOD FOR POWER-EFFICIENT CHARGING AND DISCHARGING OF A CAPACITIVE LOAD FROM A SINGLE SOURCE

Serial No. 08/986,327

Filed: December 5, 1997

Group Art Unit: 2838

Examiner: A. Berhane

SUPPLEMENT TO DECLARATION OF WILLIAM C. ATHAS (RULE 132) RE PRELIMINARY AMENDMENT (FILED JANUARY 12, 2000)

Box CPA
Assistant Commissioner for Patents
Washington, D.C. 20231

Dear Sir:

Attached is a full copy of Exhibit 2 to the Declaration of William C. Athas (Rule 132), filed January 12, 2000.

Dated: 1/13/00

Respectfully submitted,

Marc E. Brown, Reg. No. 28,590
Attorney for Applicant

OPPENHEIMER WOLFF & DONNELLY LLP
2029 Century Park East, Suite 3800
Los Angeles, CA 90067-3024
Telephone: (310) 788-5000 ◆ Facsimile: (310) 788-5100

Startup energies in energy-recovery CMOS

Sanjaya Dharmasena and Lars Svensson*
University of Southern California,
Information Sciences Institute

Nov 22 1996

Abstract

We introduce the startup energy of energy-recovery systems. We point out the inverse relationship between the startup energy and the dissipation during continuous operation. We illustrate the relationship with two CMOS circuit examples.

1 Introduction

A digital electronic circuit carries out its operations by distributing energy among its circuit nodes according to rules laid down in the interconnection of its switching devices. The energy levels on its circuit nodes—its signal energies—change as its signals transition between the logic levels used to encode the information being processed.

The energy dissipated as a result of a signal transition is, in the widely used logic design styles, at least as big as the change in the signal energy. Moreover, this switching energy frequently dominates the total dissipation of a system designed for low power. Therefore, the two directions most commonly followed in low-power design are to minimize the number of switching events, and to reduce the signal energies.

Signal energies may be reduced by decreasing the signal voltage swing [1]. This method is widely used and quite effective. It yields predictable and intuitively reasonable results due to the strong binding of switching energy to signal energy. It can, however, not be extended without bounds: most available switching devices require a minimum voltage swing on the controlling electrode to function properly. Thus, the binding of switching energy to signal energy sets a lower limit on dissipation.

Many methods and approaches have been proposed (several recently [2, 3, 4, 5]) to relax the relationship between switching energy and signal energy. The theme common to all these approaches is that signal energies inside the logic circuit are recovered rather than dissipated as heat. The

*This research is sponsored by ARPA under contract DAAL01-95-K3528.

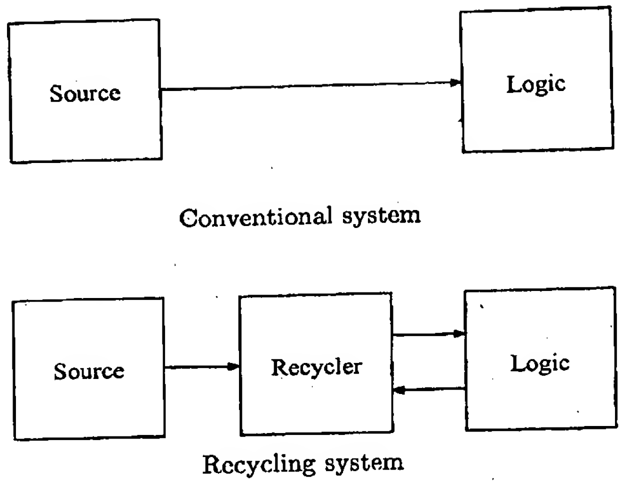

Figure 1: Conceptual conventional and energy-recovery computing systems. The arrows indicate energy flow.

concept is illustrated in Figure 1. In the conventional system, energy is delivered directly and unidirectionally from a source (such as a battery) to the logic circuit, where it is dissipated. The situation is different in the energy-recovery system. Here, energy is delivered from the source to a recycling apparatus; the latter supplies energy to and recovers energy from the logic circuit. Thus, energy transfer is unidirectional from the source to the recycler, but bidirectional between the recycler and the logic circuit.

The system organization shown in Figure 1 is conceptual only. In practical cases, it may be difficult to separate logic circuit from recycling apparatus, and recycler from source. We will discriminate source from recycler based on the direction of energy transport: a source delivers energy to the rest of the system, but never receives energy back. Furthermore, when we discuss the total energy present in an energy-recovery system, we exclude energy stored in sources; these are considered external to the system being studied.

Clearly, the energy present in an energy-recovery computing system exceeds the energy dissipated per "operation" (otherwise, there would be no energy to recover upon completion). All this energy must initially be delivered to the system to make subsequent low-power operation possible. Therefore, energy-recovery computing systems require an initial energy investment, a "startup energy," to reach the steady state where computation can be carried out efficiently. This startup energy has received considerably less attention than the energy dissipated during continuous operation. This situation is doubly unfortunate: first, for a system which does burst-like calculations and is frequently restarted, the startup energy may dominate the steady-state energy; and second, there is frequently an inverse relationship between the startup energy and the steady-state en-

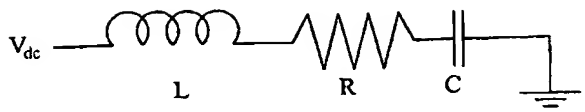

Figure 2: A series LRC circuit models a resonant energy-recovery circuit. The logic circuit contributes the resistive and capacitive components, whereas the inductance represents the power supply.

ergy, such that when a design parameter is adjusted to reduce the steady-state energy, the startup energy grows correspondingly.

The designer of an energy-recovery computing system must therefore be aware not only of how to minimize the steady-state energy, but also of how to trade it off against the startup energy. These tradeoffs are steered by relations which many designers would consider less than intuitive.

In this paper, we will discuss two examples of energy-recovery logic circuits—a resonant LRC circuit and a stepwise signal driver—and explore the relationships between startup energy and steady-state energy. Both examples show an inverse relationship between stored energy and dissipated energy. In the stepwise-driver example, we also explicitly compute the total startup energy, which shows an even stronger dependence on the dissipated energy. In both cases, we assume a CMOS technology. MOS switching devices are assumed in most discussions about low-power logic circuits: the absence of recombination of controlling and controlled charges allows not only frugal operation, but also accurate book-keeping, which simplifies evaluation of proposed circuit schemes.

2 Resonant LRC circuits

Most proposals for energy-recovery CMOS systems assume that inductors are used for temporary energy storage. We follow the examples of Koller [6] and Younis [2] and model the computing system, including the power supply, as a series LRC circuit (Figure 2). The resistive and capacitive components of the LRC circuit represent channel resistances and gate and parasitic capacitances of the CMOS logic circuit. The details of the logic style and the method of energy replenishment are invisible at this level of abstraction. Also, most proposed energy-recovery circuit styles use several clock phases, which would correspond to several resonant circuits. Given that the resonance frequency, the Q value, and the voltage swing of all phase circuits are similar, they can be adequately represented by *one* LRC circuit consisting of the parallel connections of the resistive, capacitive, and inductive components of all the clock-phase circuits.

The LRC circuit energy resonates between the capacitance and the inductance: when the inductance current is zero, the capacitance voltage is at its peak, and all the energy is stored in the capacitance; the inductance

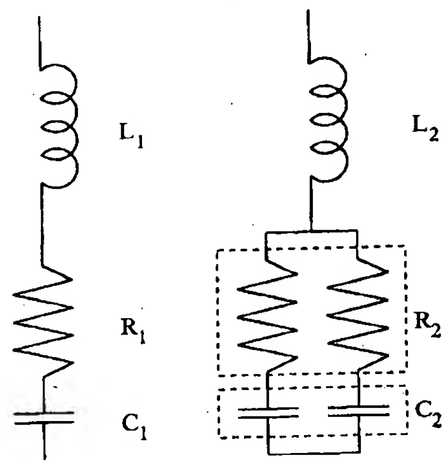

Figure 3: The model of the modified system consists of two copies of the resistive and capacitive components of the original system connected in parallel; the inductance of the modified system is selected for identical computational throughput.

current is at its peak when the capacitance is completely discharged, and all the energy is stored in the inductance. Thus, it is sufficient to know the voltage swing and the capacitance to determine the total energy present in the system.

Consider a resonant computing system modelled by an inductance L_1, a capacitance C_1, and a resistance R_1, which performs a computation in time $K_1 \cdot T_1$ (Figure 3). T_1 is the clock period, and K_1 is the number of clock cycles required to carry out the computation. Assume that $R_1 \ll \sqrt{L_1/C_1}$, so that the equivalent LRC circuit is heavily underdamped. Then:

$$T_1 \approx 2\pi\sqrt{L_1 C_1}$$

The total energy present in the model system, E_1, is determined by C_1 and the voltage swing V:

$$E_1 = \frac{1}{2} \cdot C_1 \cdot V^2$$

The energy dissipation per clock cycle is, to first order, inversely proportional to the Q value of the LRC circuit:

$$Q_1 = \frac{1}{R_1}\sqrt{\frac{L_1}{C_1}}$$

The total energy dissipation during the computation is then:

$$E_{diss1} \approx K_1 \left(\text{const} \cdot \frac{E_1}{Q_1}\right) = \text{const} \cdot K_1 E_1 R_1 \sqrt{\frac{C_1}{L_1}}$$

Next, we construct a modified system, characterized by the components L_2, C_2, and R_2, which performs a computation in time $K_2 \cdot T_2$. The modified system carries out the calculations in the same total time as the original system, i.e., $K_2 \cdot T_2 = K_1 \cdot T_1$; however, it uses increased parallelism to complete the work in half as many cycles as the original system, i.e., $K_2 = K_1/2$. Clearly, $T_2 = 2 \cdot T_1$.

For simplicity, we assume that the computations may be parallelized without overhead, such that twice the amount of computing circuitry yields twice the number of useful operations per clock cycle. Then, as illustrated in Figure 3, $C_2 = 2 \cdot C_1$ and $R_2 = R_1/2$. The required value for L_2 is given by:

$$L_2 = \frac{1}{C_2}\left(\frac{T_2}{2\pi}\right)^2 = \frac{1}{2C_1}\left(\frac{2T_1}{2\pi}\right)^2 = 2 \cdot L_1$$

Similarly, we can express the total energy, the Q-value, and the total energy dissipation for the modified system in terms of the corresponding entities for the original system:

$$E_2 = \frac{1}{2} \cdot C_2 \cdot V^2 = \frac{1}{2} \cdot 2C_1 \cdot V^2 = 2 \cdot E_1$$

$$Q_2 = \frac{1}{R_2}\sqrt{\frac{L_2}{C_2}} = \frac{2}{R_1}\sqrt{\frac{2L_1}{2C_1}} = 2 \cdot Q_1$$

$$\begin{aligned} E_{diss2} &\approx K_2\left(\text{const} \cdot \frac{E_2}{Q_2}\right) = \frac{K_1}{2}\left(\text{const} \cdot \frac{2E_1}{2Q_1}\right) \\ &= \frac{1}{2} \cdot E_{diss1} \end{aligned}$$

Thus, the total dissipation associated with the calculation has been reduced by a factor of 2. The total energy present in the computing system, however, has increased by the same factor.

The energy present in the resonant system during operation constitutes a lower limit on the energy injected into the system during startup. Some of the startup energy will not remain in the system, but rather be dissipated in the process of moving the system to its low-power state. This overhead depends highly on the exact circuit solution; no estimate is offered here.

3 Stepwise driver circuits

Unlike the LRC circuit, a stepwise driver does not require any additional circuits for startup. Therefore, its operation can be modelled more accurately than that for the resonant circuit. Furthermore, a single analysis yields both the startup cost and the cost per "operation".

Consider a stepwise driver with N steps, consisting of a capacitive load, C_L, and $N-1$ tank capacitors with identical capacitance, C_T, as shown in Figure 4. The tank capacitors play the role of the recycler in this configuration.

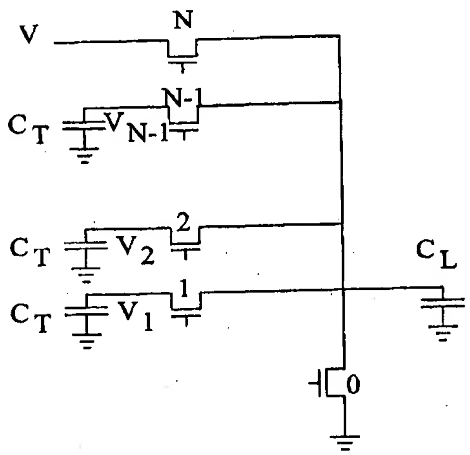

Figure 4: A stepwise signal driver uses a bank of tank capacitors for temporary energy storage.

Let V_i be the voltage of the ith tank capacitor. The load is charged by briefly connecting V_1 through V_N in succession (by opening switch 0 and closing switch 1, waiting for the load to reach V_1, opening switch 1 and closing switch 2, etc.). Switch N is kept closed for as long as the load capacitance is to stay charged. The load is discharged by connecting it to V_{N-1} through V_1, and finally to ground by closing switch 0.

We assume that the circuit components are linear and that each MOS switch is turned on for a duration sufficiently long enough to complete the charging/discharging of the load C_L. We also assume that the tank capacitance C_T is much larger than the load capacitance C_L. With these assumptions, the following result can be shown.

Theorem 3.1 *Assume that a N stepwise driver circuit operates under the previously stated assumptions. Then, the tank capacitor voltages converge to the following values:*

$$V_i = \frac{i}{N} V \text{ for } i = 1, 2, \ldots, N-1$$

The convergence rate R is given by:

$$R = \max\{|\lambda_{max}|, |\lambda_{min}|\}$$

Here, λ_{max} and λ_{min} are given by:

$$\begin{aligned} \lambda_{max} &= 1 - 2\frac{C_L}{C_T} + 2\frac{C_L}{C_T}\cos\left(\frac{\pi}{N}\right) \\ \lambda_{min} &= 1 - 2\frac{C_L}{C_T} - 2\frac{C_L}{C_T}\cos\left(\frac{\pi}{N}\right) \end{aligned}$$

Proof: See Appendix A for details.

As for the resonant circuit, we can explore the relationships between startup energy and dissipated energy. For each "step" from V_i to V_{i+1}, the dissipation is given by the transferred charge and the average voltage drop across the switch:

$$E_{diss,step} = Q\bar{V} = C_L\frac{V}{N}\cdot\frac{V}{2N} = \frac{1}{2}C_L\frac{V^2}{N^2}$$

To charge C_L all the way to the supply voltage V, N steps are used. The dissipation for a full transition from 0 to V is then:

$$E_{diss,N-step} = N\cdot E_{step} = \frac{1}{2}C_L\frac{V^2}{N} \qquad (1)$$

This formulation for the energy dissipation ignores the energy needed to control the MOS switches. It also assumes that the tank capacitor voltages do not undergo any fluctuations during a complete charge-discharge cycle. If this assumption is removed, then equation (1) must be modified to include a correction term which reflects an increased dissipation on account of such fluctuations.

The energy stored in the stepwise driver after convergence is:

$$\begin{aligned} E_{stored} &= \frac{1}{2}C_T\sum_{i=1}^{N-1}\left(\frac{i\cdot V}{N}\right)^2 \\ &= \frac{1}{2}\frac{(N-1)(2N-1)}{6N}\cdot C_TV^2 \end{aligned} \qquad (2)$$

We now investigate how the energy dissipation $E_{diss,N-step}$ and the stored energy E_{stored} depend on the stepwise driver parameter N.

First, consider the case when N is increasing, but C_T is constant. The stored energy E_{stored} increases linearly with N. Since dissipation is inversely proportional to N, $E_{diss,N-step}$ decreases linearly with N. This behavior is analogous to that exhibited by the LRC resonant circuit.

Next, consider the case when N is increasing, but the sum of all the tank capacitances, C_{total}, remains constant. In this case, the tank capacitance decreases with N:

$$C_T = \frac{C_{total}}{(N-1)} \qquad (3)$$

The stored-energy equation (2) can be rewritten as:

$$E_{stored} = \frac{1}{2}\frac{(2N-1)}{6N}\cdot C_{total}V^2 \qquad (4)$$

Hence, from equation (4) it follows that the stored energy remains constant (for large N) despite an increasing N. However, the dissipation energy falls with N. Note that N cannot be increased arbitrarily beyond a certain limit to reap the benefit from a diminishing dissipation. Increasing N brings a reduction in the per-step tank capacitance C_T given by (3). Therefore, there exists a limit for N at which the per-step tank capacitance C_T value becomes comparable to the load capacitance C_L.

At this point, the assumption made regarding the stepwise driver convergence no longer holds. No claims can be made regarding the convergence in this case, since the mathematical model is not representative of the actual circuit behavior.

We can also derive an approximate expression for the energy required to bring the stepwise-charging system to a state where all the tank capacitor voltages have converged to their steady-state values:

$$E_{startup} \approx \frac{N(N-1)}{\pi^2} \cdot C_T V^2 \tag{5}$$

See Appendix B for derivation of this formula. Interestingly, this expression grows as N^2, whereas the stored energy grows as N. In this system, a dissipation reduction therefore requires a linear increase in system energy, but a *quadratic* increase in the startup energy.

The dependence can be brought back to linear by allowing more than one parameter to vary. With a constant *total* tank capacitance, such that $C_T = C_{total}/(N-1)$, the startup energy is linear in N. However, our derivations are not valid unless $C_L \ll C_T$, so such a system is not scalable to large N.

We finally observe that for large N, the ratio of the stored energy to the startup energy is proportional to that of the dissipated energy and the signal energy $E_0 = C_L V^2$:

$$\frac{E_{stored}}{E_{startup}} \sim \frac{E_{diss,N-step}}{E_0}$$

This expression illustrates the problem of optimizing the efficiency of both the initialization and the operation of the stepwise driver.

4 Conclusion

We have examined two different energy-recovery systems, using very dissimilar approaches. In both cases, the energy stored in the system increases linearly with decreasing per-operation dissipation. This similarity is remarkable in view of the different approaches: in one case, we varied the amount of computational parallelism of the system, whereas in the other case, we changed the granularity of the charging.

If the energy stored in the system is discarded, i.e. dissipated, when the computations are completed, the per-operation dissipation cannot usefully be reduced below the point where the system energy dominates the overall dissipation. In such cases, the overall energy cost can be reduced by allowing the per-operation dissipation to *increase*. This tradeoff is most important for systems which do only brief computations separated by long periods of dormancy.

Counter-intuitive tradeoffs such as this one also occur elsewhere in the design of energy-recovery circuits. A prime example is the choice of voltage swing in transmission-gate-based logic: reducing the voltage swing below $4V_{th}$ results in increased dissipation, even though the signal energy is reduced [4].

In summary, energy-recovery circuits display complex relationships and tradeoffs between signal energy, switching energy, system energy, and startup energy. Of these, the signal energy and switching energy has received most attention, but the other relationships may be as important for the design of truly efficient energy-recovery systems.

5 Acknowledgement

The authors would like to thank Dr. J. Koller for the initial remarks which provided the inspiration for this paper; and Dr. W. Athas for helpful discussions.

References

[1] CHANDRAKASAN, A. P., S. SHENG, and R. W. BRODERSEN, "Low-power CMOS digital design", *IEEE Journal of Solid-State Circuits*, vol. 27, no. 4, (April 1992), 473-484.

[2] YOUNIS, S. G., and T. F. KNIGHT, "Practical implementation of charge recovery asymptotically zero power CMOS", in *Proc. 1993 Symp. on Integrated Syst.*, MIT Press, (1993), 234-250.

[3] DICKINSON Alex G., and John S. DENKER, "Adiabatic dynamic logic", *IEEE Journal of Solid-State Circuits*, vol. 30, no. 3, (March 1995), 311-315.

[4] ATHAS, William C., Lars "J." SVENSSON, Jeffrey G. KOLLER, Nestoras TZARTZANIS, and Eric Ying-Chin CHOU, "Low-power digital systems based on adiabatic-switching principles", *IEEE Transactions on VLSI Systems*, vol. 2, no. 4, (December 1994), 398-407.

[5] SVENSSON, Lars "J.", and Jeffrey KOLLER, "Driving a capacitive load without dissipating fCV^2", in *1994 IEEE Symposium on Power Electronics Digest of Technical Papers*, IEEE Solid-State Circuits Council, (1994), 100-101.

[6] KOLLER, Jeffrey G., and William C. ATHAS, "Adiabatic switching, low energy computing, and the physics of storing and erasing information", in *Proc. Workshop on Physics and Computation*, PhysCmp '92, (Oct. 1992); IEEE Press, (1993).

A Convergence rate calculation

Since the tank capacitance, C_T, is assumed to be much larger than the load capacitance C_L, the voltage on each tank capacitor can be assumed to remain constant during a charge-discharge cycle. Hence, the charge pulled from tank capacitor i during charging in cycle t is:

$$q_{i,up}(t) = C_L\left(V_i(t) - V_{i-1}(t)\right).$$

The charge pushed back into tank capacitor i during discharging is:

$$q_{i,down}(t) = C_L\left(V_{i+1}(t) - V_i(t)\right).$$

Hence, the net increase on tank capacitor i over the charge-discharge cycle t is:

$$\begin{aligned} q_i(t) &= q_{i,down}(t) - q_{i,up}(t) \\ &= C_L\left(V_{i+1}(t) + V_{i-1}(t) - 2V_i(t)\right). \end{aligned}$$

The corresponding voltage change on tank capacitor i is:

$$\Delta V_i(t) = \frac{C_L}{C_T}\left(V_{i+1}(t) + V_{i-1}(t) - 2V_i(t)\right). \tag{6}$$

Thus, the voltage after the charge-discharge cycle is:

$$V_i(t+1) = V_i(t) + \Delta V_i(t). \tag{7}$$

We introduce the voltage deviances, $\tilde{V}_i$, $i = 1, 2, \ldots, N-1$, as the differences between the actual tank voltages and the voltages given by the even distribution:

$$\tilde{V}_i(t) \triangleq V_i(t) - \frac{i}{N}V. \tag{8}$$

For notational convenience, let $\tilde{V}_0(t) = \tilde{V}_N(t) \equiv 0\ \forall t$. Rewrite equation (6) in terms of the deviances:

$$\begin{aligned} \Delta\tilde{V}_i(t) &= \Delta V_i(t) \\ &= \frac{C_L}{C_T}\left(\tilde{V}_{i+1}(t) + \tilde{V}_{i-1}(t) - 2\tilde{V}_i(t)\right) \end{aligned} \tag{9}$$

Then, combining equations (7) – (9), we obtain a linear system of equations:

$$\tilde{\mathbf{V}}(\mathbf{t}+1) = \mathbf{G}\cdot\tilde{\mathbf{V}}(\mathbf{t}) \tag{10}$$

The variables are defined as follows:

$$\tilde{\mathbf{V}}(\mathbf{t}+1) = \left[\tilde{V}_1(t+1), \tilde{V}_2(t+1), \ldots, \tilde{V}_{N-1}(t+1)\right]^T$$

$$\mathbf{G} = \begin{bmatrix} 1-2K & K & 0 & \cdots & 0 \\ K & 1-2K & K & \cdots & 0 \\ 0 & K & 1-2K & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1-2K \end{bmatrix} \tag{11}$$

where $K = \frac{C_L}{C_T}$. It is possible to show that the following inequality holds:

$$\|\tilde{V}_i(t+1)\| \leq R \cdot \|\tilde{V}_i(t)\|$$

R is related to the maximum and the minimum eigenvalues of the matrix $\mathbf{G}$ defined in (11):

$$R = \max\{|\lambda_{max}|, |\lambda_{min}|\}$$

$$\begin{aligned}\lambda_{max} &= 1 - 2\frac{C_L}{C_T} + 2\frac{C_L}{C_T}\cos\left(\frac{\pi}{N}\right)\\ \lambda_{min} &= 1 - 2\frac{C_L}{C_T} - 2\frac{C_L}{C_T}\cos\left(\frac{\pi}{N}\right)\end{aligned}$$

R represents the rate of evolution of the error vector $\tilde{V}_i(t)$ and must be constrained to be less than unity for convergence. This is possible if and only if

$$1 - 2\frac{C_L}{C_T} - 2\frac{C_L}{C_T}\cos\left(\frac{\pi}{N}\right) > -1$$

from which it follows that

$$C_T > \left[1 + \cos\left(\frac{\pi}{N}\right)\right] C_L \tag{12}$$

is sufficient for convergence.

It must be noted that the lower bound for the tank capacitance given by (12) is sufficient but not necessary for convergence. If C_T is less than this bound, nothing can be said of convergence since the validity of the mathematical model holds only when C_T is much larger than C_L.

B Startup energy

The energy invested in the stepwise driver to bring the tank voltages to converge is injected from the power supply at voltage V. Assume that all tank capacitor voltages are initially 0. In each charge-discharge cycle, the amount of charge injected is given by $Q_{inj} = (V - V_{N-1})C_L$.

When the tank voltages have converged, $V - V_{N-1} = \frac{1}{N}V$; the larger voltage difference during startup cause auxiliary charge to be drawn from the supply. The total auxiliary charge delivered by the supply, $Q_{startup}$, can be approximated by:

$$Q_{startup} = \sum_{k=0}^{\infty}\left[\left(\frac{N-1}{N}\right) - \left(1-\lambda^k\right)\left(\frac{N-1}{N}\right)\right] C_L V \tag{13}$$

where $\lambda = 1 - 2\frac{C_L}{C_T} + 2\frac{C_L}{C_T}\cos\left(\frac{\pi}{N}\right)$.

Here, λ is chosen as the slowest eigenvalue of the matrix defined in (11) with C_T assumed to be at least twice as large as C_L. The motivation for the choice of λ is that the speed of convergence is influenced by the slowest eigenvalue. Although use of this eigenvalue in equation (13) leads to a somewhat conservative result, it is adequate for our analysis, and furthermore, it simplifies the formulation considerably.

Simplifying equation (13), we get:

$$\begin{aligned}Q_{startup} &= \frac{N-1}{N}\cdot\frac{1}{1-\lambda}\cdot C_L V\\ &= \frac{N-1}{N\left[2\frac{C_L}{C_T} - 2\frac{C_L}{C_T}\cos\left(\frac{\pi}{N}\right)\right]}\cdot C_L V\\ &\approx \frac{N(N-1)}{\pi^2}\cdot C_T V\end{aligned}$$

for sufficiently large N

So,

$$\begin{aligned} E_{startup} &= Q_{startup} \cdot V \\ &\approx \frac{N(N-1)}{\pi^2} \cdot C_T V^2 \end{aligned}$$